402134

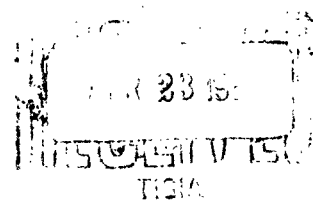# TOWARDS A REVIVAL OF THE STATISTICAL LAW OF PARETO
(Second Draft)*

by

Benoit Mandelbrot**

3/22/62

ABSTRACT: This is a defense and illustration of the statistical
law of Pareto, and an informal introduction to its role--and that of
certain of its kins--in the study of price variations, of income dis-
tributions, of the distributions of the sizes of firms and cities, and
of related questions in economics and in other social and physical
sciences.

IBM

This work, submitted for publication to <u>The Journal of Political Economy</u>, is a new draft of a previously circulated preprint, and it also incorporates a revised version of my IBM Note "Aggregation, Choice, Mixture and the law of Pareto." To avoid the possibility of obsolete reference or of any other confusion, please kindly discard the copies, which you may have in your possession, of either the earlier draft or the earlier Note.

1. Introduction.

Neglect and even contempt often mark the attitude of statisticians and of mathematical economists towards Pareto's well-known empirical discovery, that there exist two constants C and $\alpha > 0$, such that the relative number of incomes exceeding u can— for large values of u— be written in the form $Cu^{-\alpha}$ (footnote 1).

It is not very seriously questioned, however, that the law of Pareto represents very satisfactor y, not only the "tail" of the distribution of personal income, but also those of the distribution of firm sizes and of city sizes. In fact, the game consisting of searching for new instances of that law has been at times very popular and quite successful, although seldom respected [see for example the writings of George Kingsley Zipf (17) (18)].

We think therefore that the law of Pareto has been neglected because it does ot represent the middle range of incomes-- which may be the more important for certain purposes-- and also because it is so lacking of theoretical motivation-- at least within the context of elementary probability theory. We believe, however, that in the light of modern advances in the pure theory of random variables and of stochastic processes, this remarkable finding deserves a systematic new examination.

We shall see indeed that the law of Pareto <u>literally thrusts itself</u> upon anyone who takes seriously the models of economics based on maximization or upon linear aggregation, upon anyone who takes a cautious view of the origin of the economic data, and upon anyone who believes in the influence on economics of the physical distribution of various scarce natural resources.

We shall also show the following: when the "spontaneous activity" of a system is ruled by a Paretian process, the causally structural features of the system are likely to be very much more hidden by noise than is the case where the noise is Gaussian. In fact, causal structures may be totally "drowned out." On the other hand, Paretian noise generates all kinds of "patterns" that seem to be perfectly clear-cut but have no value for purposes of prediction. Thus, in the presence of a Paretian "spontaneous activity," the scientist is faced by an unexpectedly heavy burden of proof, and the basic problem of the validation of laws acquires many new and indeed perturbing features.

We shall see that the most important features of the law of Pareto are linked to the length of its tail, and not to its extreme skewness. In fact, in the cases where we shall deal with random variables that can also take large negative values, we shall have to introduce a family of bilateral Paretian distributions, which may even be symmetric. Hence, the extreme skewness of the distribution of income must be considered as being a secondary feature of those Paretian variables that are constrained to be positive.

The general "tone" of this paper is indicated by its title. We shall not attmept to treat any point exhaustively, nor to fully specify all the conditions of validity of our assertions, which are discussed in detail in the publications referred to in the bibliography.

The last section of this paper will examine two of the most influential critiques of the law of Pareto.

2. The general principle of our "method of invariant laws."

The approach used in our studies of the law of Pareto may seem unusual in the context of social science, but it resembles a method very familiar in physics (footnote 2). To begin with, we find that the various "microscopic models," which could be considered as explaining "why" such and such a version of the law of Pareto is encountered in such and such a domain, are at the very best hardly more convincing than the law itself, because they are of much less general applicability, and because seemingly slight and irrelevant changes in the hypotheses completely change the result. Moreover, we believe that the stress upon generative models of the law of Pareto has handicapped the study of its remarkable properties.

Therefore, we have preferred to center our work in this area around the study of the actual conditions of empirical observation, as practiced in economics and in other social sciences. By "observation" we not only mean the activity of the scholar who observes to describe, but also that of the entrepreneur who observes to act. In both cases, we note that, even if irreducible economic quantities had a real existence, they could hardly ever be observed directly; they would rather be altered by some ill-known sequence of all kinds of manipulation.

In most practical problems, very little can be done about this difficulty, and one must make do with whatever approximation to the desired data is

actually available. But inappropriate data are a notorious handicap in theoretical work, since economic relationships are usually relative to conceptual irreducible economic quantities, and cannot generally be expected to be left invariant by the manipulations performed before actual measurement. That is, the analytical formulas, by which they may be described, must be expected to change in form markedly, whenever one applies one of the basic transformations. As a result, however great the practical importance of these relationships, and hence however great the efforts to understand them, there is a good chance that their form will be discovered later, and that they will forever remain known with lesser precision, than the phenomena that "happen" to be in some sense invariant with respect to the maximum number of observational transformations, such as the following (which are all funda-mental, but unequally so).

Linear aggregation, or simple addition of various quantities in their common natural scale. For example, aggregates of various kinds of income are better known than each kind taken separately. Long-term changes in most economic quantities are better known than the more desirable medium-term changes; moreover, the meaning of "medium-term" differs between series, so that a law that is not invariant under aggregation would be apparent in some series, and not in others, and could not be firmly established. A number of operations of aggregation also occur in the context of firm sizes, in particular when "old" firms merge within a "new" one.

The most universal interpretation of aggregation occurs, however, in linear models that add the (weighted) contributions of several "causes", or more generally embody linear relationships between several variables or between the current and the past values of a single variable (autoregressive schemes). The scholar's preference for such models is of course based upon the unhappy but unquestionable fact that mathematics offers few workable non-linear tools to the scientist.

There is clearly nothing new in our emphasis upon invariance under aggregation. It is indeed well known that the sum of two independent Gaussian variables is itself Gaussian, and -- after the ease of analytical manipulation -- this is the principal reason for using Gaussian "error terms" in linear models. However, the Gaussian law is alone to be invariant under aggregation only if one excludes random variables with infinite

populations moments (whereas we shall not exclude them; see section 5).
(Besides, the Gaussian law is not invariant under our other two observational transformations).

Let us also note that one may aggregate a small or a very large number of quantities. Whenever possible, "very large" is approximated by "infinite", so that aggregation is intimately related with the question of the central limit theorem concerning the behavior of limits of sums of random variables.

A second fundamental transformation is weighted mixture, or compounding. For example, a compounded lottery ticket would be one in which a first preliminary chance drawing would determine in which of several final drawings the gambler has the right to participate. This provides a model for all kinds of actually observed variables: For example, if one does not know the precise origin of a given set of income data, one may consider that they were picked at random among a number of possible basic distributions; the distribution of observed incomes would then be the mixture of the basic distributions. Similarly, price data often refer to grades of a commodity that are not precisely known and hence can be assumed to be randomly determined. Finally, the notion of a firm is somewhat undeterminate (what about almost wholly owned, but legally distinct subsidiaries?), and available data refer to firms that may vary in size between individual establishments and holding companies; such mixture may be represented by random compounding.

In many cases, one has to deal with a combination of the above operations: for example, after a wave of mergers has hit an industry, one may consider that the distribution of "new" firms is the mixture of the distribution of companies not involved in a merger, of the distribution of companies made up of the sum of two old firms, and perhaps even of sums of more than two firms.

The final basic transformation is maximizing choice, i.e., the selection of the largest or smallest quantity in a set. For example, it may be that all we know about a set of quantities is the size of the one chosen by a profit-maximizer. If one must use historical data, one must often expect to find that only the exceptional largest or smallest events are fully reported, for example, droughts or floods, famines (and the names of the "Bad Kings" who reigned in those times), or "Good times" (and the names of the "Good Kings"). Mixture and maximization are often mixed, since many data are a mixture of fully reported periods and of reporting limited to the extreme cases.

Although the above transformations are not the only ones of interest, they are so important, that one must characterize the laws which they leave invariant. It so happens, that invariance-up-to-scale holds asymptotically for all three transformations if the parts follow the law of Pareto (in the case of infinite aggregation, invariance only holds if Pareto's exponent is less than two). On the contrary (with some qualifications) invariance does not hold -- even asymptotically -- in any other case. Hence, if one's belief in the importance of those transformations has any strength at all, one will attach a special importance to Paretian phenomena, at least from a purely pragmatic viewpoint.

This also affects the proper presentation of empirical results: Indeed, one knows that, in order to be precise in the statement of scientific laws, it is not sufficient to say that income, for example, is Paretian; it is also necessary to list the excluded alternatives. Our considerations will show that the proper precise statement is not of the form: "it is true that incomes (or firm sizes) follow the law of Pareto; it is not true that incomes follow either the Gaussian, or the Poisson, or the negative binomial or the log-normal law." We must rather say: "it is true that incomes (or firm sizes) follow the law of Pareto; it is not true that the distributions of income are very sensitive to the methods of reporting and of observation."

3. Some invariance properties of Pareto's law and of certain of its kins.

Of course, the singular character of the asymptotic law of Pareto holds only under additional assumptions, so that the problem will surely not be exhausted by our present approach. We shall, indeed, consider N independent random variables, U_n $(1 \leqslant n \leqslant N)$, following the weak (asymptotic) form of the law of Pareto, with the same exponent α:

$$\Pr(U_n \geqslant u) \sim C_n u^{-\alpha} \text{ if } u \text{ is large.}$$

The behavior of $\Pr(U_n < -u)$ for large u will be examined in section 7.

Keeping the proofs in footnotes, we shall begin by quoting some statements that imply that a Paretian behavior of U_n is sufficient for the three types of asymptotic invariance -- up-to-scale. The sign $\sum^N$ will always refer to the addition of the terms relative to the/possible values of the index n.

Weighted Mixture. Suppose that the random variable U_W is a weighted mixture of the U_n, i.e. that it has the probability p_n of being identical to U_n. One can show (footnote 3) that this U_W is also asymptotically Paretian, and that its scale parameter is $C_W = \sum p_n C_n$, which is the __weighted average__ of the separate scale coefficients C_n.

Maximizing choice. Let U_M be the largest of the variables U_n, (that is, the one that turns a posteriori to be the largest, when the values of all the U_n are known; there is no simple way of saying which one of N random variables is the largest!). One can show (footnote 4) that this U_M is also asymptotically Paretian, with a scale parameter which is the __sum__ of the separate scale coefficients C_n.

Aggregation. Let C_A be the sum of the random variables U_n. One can show (footnote 5   ) that it is also asymptotically Paretian, with a scale parameter that is again the __sum__ of the separate C_n. Thus, the sum of the U_n behaves asymptotically exactly like the largest of them.

Mixture combined with aggregation -- an operation that occurs in the theory of mergers -- also leaves the law of Pareto invariant up to scale.

The converse of the above statements are true only in the first approximation: in order for the invariances-up-to-scale to hold, the distributions of the U_n need not strictly follow the law of Pareto; but the actual generalizations are in practice quite negligible.

Strictly invariant and limit distributions.

Let us now abandon asymptotics and let us introduce Fréchet's and Lévy's kins of the law of Pareto, by imitating (with a different interpretation), a famous principle of physics: to require that the random variables U_n be strictly invariant -- (up-to-scale) with respect to one of our three transformations. This means the following: let the N random variables U_n all follow--up to changes of scale--the same law as the variable U, so that U_n can be written as $a_n U$, where $a_n > 0$; we shall require that U_W (respectively U_M or U_A) also follow--up to scale--the same law as U. For that, it must be possible to write U_W (respectively U_M or U_A) in the form $a_W U$ (respectively $a_M U$ or $a_A U$) is some positive function of the numbers a_n.

It turns out that the conditions of invariance lead to somewhat similar equations in all three cases (see footnote 6   ). More precisely, one obtains the following results:

Maximization. The invariant laws must be of the form $F_M(u)=\exp(-u^{-\alpha})$, which is due to Maurice Fréchet (reference 5). They are clearly Paretian, since-- for large u-- F_M can be approximated by $1 - Cu^{-\alpha}$. They also "happen" to have the remarkable property of being the limit distributions of the expression $N^{-1/\alpha} \max U_n$, where the U_n are asymptotically Paretian. There are no other distributions that can be obtained simply by multiplying max U_n by an appropriate factor and by having N tend to infinity. (If one also allows the origin of U to change as $N \to \infty$, one can obtain the Fisher-Tippett distribution, which is not Paretian and is not invariant under the other two transformations.)

Mixture. In this case, invariance leads to $F_W(u)=1-Cu^{-\alpha}$, i.e., to the law of Pareto extended down to u = 0, an expression which corresponds to an infinite total probability. One notes immediately that such a solution is strictly speaking inacceptable. However, it must not be rejected offhand, because in many cases in practice U is further restricted by some relation of the form $C < a \leq u \leq b$, leading to a perfectly acceptable conditional probability distribution.

Aggregation. Finally, aggregation leads to random variables that are part of the family of Lévy's "stable distributions," other members of which will be encountered later. (See reference 4) One knows $dF_A(u)$ in closed form for the stable law with $\alpha = 2$ (which is the Gaussian in a sense; it is a limit case of the other stable Paretian laws, but is not itself Paretian) that with $\alpha = 1/2$, which plays a central role in return to equilibrium in coin tossing. Otherwise, no closed analytic expression is known for the stable $F_A(u)$; Lévy has shown, however, that unless $\alpha = 2$, they asymptotically follow the law of Pareto of exponent α .

The stable variables yielded by the present argument can take negative values if $1 \leq \alpha \leq 2$, as is readily seen in the Gaussian case. But the probability of large negative values is very small, and we have shown in our papers how to handle this question in practice, with the help of appropriate changes of origin.

Lévy's stable distributions have another important property: they are the only possible non-Gaussin limits of linearly weighted sums of random variables. Hence, even though they cannot begin to compare with the Gaussian law from the viewpoint of ease of mathematical manipulation, they share both

the fundamental properties of that law from the viewpoint of linear oper-
ations: the existence of the corresponding forms of the non-classical cen-
tral limit theorem show that, if a process is the resultant of many additive
contributions, it need not be Gaussian; if one wishes to explain by linear
addition a phenomenon that is ruled by a skew distribution, it is not
necessary to assume that the addition in question is performed in the scale
of log U rather than in the scale of U itself. This also shows that the
lognormal distribution is not the only skew law that can be explained by addi-
tion arguments; this takes away the principal asset of that law, which is
known in most cases to grossly underestimate the largest values that can be
taken by the variable of interest.

One can see that the probability densities of the three invariant
families differ through most of the range of u. However, if $0 < \alpha < 2$,
their asymptotical behaviors coincide, so that the law of Pareto is also
asymptotically invariant with respect to applications of an arbitrary
succession of the basic transformations.

It should be noted that Fréchet's and Lévy's Paretian limit distri-
butions have attracted substantial attention from pure mathematicians. How-
ever, the generally known applications of Paretian maximum distributions were
few and those of Paretian sum distribution (stable laws) were practically
non-existent. It is true that the introduction of the Gnedenko-Kolmogoroff
treatise (reference 4) contains statements about the wide applicability
of the mathematical techniques to which that book is devoted, and even
references to forthcoming publications specially concerned with applications.
However, when we discussed this introduction with the senior author in
1958 (ten years after the appearance of the original Russian book), we
found that these papers had not materialized after all — for lack of
applications. Basically, the only fairly well-known practical instance
of a stable distribution remains the law, due to Holtsmark but often re-
discovered, that rules the Newtonian attraction between randomly distri-
buted stars (see reference 7). Anyway, our plea, that stable laws be
counted among the most "common" probability distributions, has not been
made void by the Gnedenko-Kolmogoroff book.

4. On the value of the evidence of doubly logarithmic graphs.

The above limitation in the value of α brings us to another,
quite different, aspect of the general problem of observation, relative to
the practical significance of statements having only an asymptotic valid-
ity. Indeed, in order to verify empirically the law of Pareto, the usual
first step is to draw the so-called doubly logarithmic graph of $\log_{10}[1 - F(u)]$
as a function of $\log_{10}u$. One should find that this graph is a straight line
with the slope $-\alpha$, or at least that it rapidly becomes straight with this
slope. But let us look closer at the empirical point of largest u. Ex-
cept for the distributions of incomes, one has at most a sample of 1000 or
2000 items; or one may otherwise know the value of u that is exceeded with
the frequency $1 - F(u) = 1000^{-1}$ or 2000^{-1}. That is, the "height" of the
empirical doubly logarithmic graph will at the very best cover <u>three</u> units
of the decimal logarithm of 1 - F. The "width" of this graph will therefore
be at the very best equal to $3/\alpha$ units of the decimal logarithm of u. How-
ever if one wants to estimate reliably the value of the slope α , it is
necessary that the width of the graph be close to one unit: therefore, one
cannot have any trust whatsoever in data that suggest that α is larger than
3, and the practical range of alphas is anyway hardly wider than in the case
of Lévy's Paretian laws.

Looking at the same question from another angle, let us plot a Guassian,
lognormal negative binomial or exponential distribution on doubly logarithmic
paper: since these distributions are all very "short-tailed," the slope of
the graph will become asymptotically infinite. However, in the region of
probabilities down to 1000^{-1}, the dispersion of empirical data is liable to
generate-- on doubly logarithmic coordinates-- the appearance of a straight
line having a high but finite slope. In the words of F. Macaulay (see
section 9): "The approximate linearity of the tail of a frequency distribu-
tion charted on a double logarithmic scale signifies relatively little, be-
cause it is such a common characteristic of frequency distributions of many
and various types." However, linearity with a low slope signifies a great
deal indeed. (see Figure 1)

There is another way of describing curve-fitting using special papers:
one may say that the maximum distance between the sample curve and some re-
ference curve-- preferably a straight line-- defines a kind of distance

between two alternative probability laws. Any special paper-- whether it be lognormal or Paretian-- should be used only in ranges where the distances which it defines are sensitive to the differences that count from the viewpoint of the problems at hand. Hence, the conservative thing to do is often to consider several hypotheses, i.e., to use several kinds of paper.

To sum up, if one takes account of mixtures, maximization and practical measurement the range of values of alpha is reduced to the interval from 0 to 3. If one also takes account of aggregation, α must fall between 0 and 2 (actually, the range of "apparent" alphas is somewhat wider).

5. The problem of the meaning of random variables with infinite population moments.

Such Paretian laws are extraordinarily long-tailed, as measured by Gaussian standards. In particular, if $\alpha < 2$, the population second moment is infinite. It should be stressed, however, that there is nothing improper in such a notion.

It is of course true that-- observed variables being finite-- the sample moments of all orders are themselves finite for finite sample sizes; but this does not exclude that they become infinite with the sample size. It may also be true that the asymptotic behavior of samples is practically irrelevant, because the sizes of all empirical samples are by nature finite. For example, one may argue that the history of cotton prices is a finite set of data from 1816 to 1958, because speculation on cotton was very much diminished by the 1958 acts of the Congress of the United States. Similarly, when one studies the sizes of United States cities, one deals with statistical populations for which the sample size is bounded. Even for continuing series, one may well argue for "après moi, le Déluge," and neglect any time horizon longer than a man's life. Hence, the behavior of the moments for infinite sample sizes may seem unimportant. But all that this actually implies is that the only meaningful consequences of infinite population moments are those relative to the sample moments of increasing sub-sets of our various bounded universes. Here, the situation is basically as follows: (see Figures 2 and 3)

There is no question that, wherever the sample second moment is ob-

served to rapidly "stabilize" around the value corresponding to the total
set, it is useful to take that value as an estimate of the population
second moment of a conjectural infinite population, from which the sample
could have been drawn. But suppose that the sample second moments corres-
ponding to increasing sub-sets continue to vary widely, even when the
sample size approaches the maximum imposed by the subject matter. From
the viewpoint of sampling, this must be interpreted as meaning that the dis-
tribution is such that even the largest available sample is too small for
reliable estimation of the population second moment, or-- in other words--
that a wide range of values of the population second moment are equally
compatible with the data. Moreover, it frequently turns out that this range
of values of the moment happens to include the value "infinity," implying
that facts can be equally well described by assuming that the "actual" mo-
ment is extremely large but finite, or by assuming that it is infinite.

In order to motivate the alternative that we prefer, let us point
out that a realistic, scientific model must not depend too critically upon
quantities that are difficult to measure. The finite-moment model is un-
fortunately very sensitive to the value of the population second moment,
and there are many other ways in which the first assumption, which of course
is the more reasonable a priori, also happens to be by far the more cum-
bersome analytically. The second assumption on the contrary leads to simple
analytical developments, and the rapidity of growth of the sample second
moment can be so adjusted that it would lead to absurd results only if one
applied it to "infinite" samples, that is, if one raised problems devoid of
concrete meaning.

In other words, there is no danger in assuming, as we shall do, that
an intrinsically bounded variable is drawn at random from an infinite popu-
lation of of unbounded variables having an infinite second moment. But all
those infinities are a relative matter, entirely dependent upon the statis-
ticians' span of interest; as the maximum useful sample size increases the
range of the estimates of the second moment will steadily narrow. Hence,
beyond a limit, the second moments of some variables may have to be con-
sidered as actually being finite; conversely, there are variables for which
the second moment must be considered as being finite only if the useful
sample size is shorter than some limit.

Actually, our use of infinity is a most common one in statistics,

insofar as it concerns the function $\max(u_1, u_2, \ldots u_N)$ of the observations.
From this viewpoint, even the use of infinite spans would seem to be im-
proper; however, it is well known in statistics that little could be done
if one could not use unbounded variables: one even uses the Gaussian to
represent the height of adult humans, which is surely positive!

The unusual behavior of the moments of Paretian distributions can be
used to introduce the least precise interpretation of the validity of the
law of Pareto. For example, if the first moment is finite, but the second
moment is infinite, the function $1 - F(u)$ must decrease slower than $1/u^2$
but faster than $1/u$ when u tends to infinity. In this case, the behavior of
$F(u)$ in the tails is very important, and, in the first approximation, it
may be very useful to approximate it by the form $Cu^{-\alpha}$, with $1 < \alpha < 2$;
this can never lead to harm, as long as one limits oneself to consequences
that are not very sensitive to the actual value of α. If on the contrary
the tail is very short (say if moments are finite up to the fourth order)
the behavior of the function $F(u)$ for large u is far less important to re-
present than its behavior elsewhere; hence, one will risk little harm with
interpolations by the Gaussian or the lognormal distribution.

6. Problems of statistical inference and of confirmation of scientific laws, when the "background noise" is Paretian.

It is well known that second moments are heavily used in statistical
measures of dispersion or of "standard deviation." Hence, whenever the
considerations of section 5 are required to explain the erratic behavior
of sample second moments, a substantial portion of the usual methods of sta-
tistics should be expected to fail, except if extraordinary care is exerted.
Examples of such failure have of course often been observed empirically, and
have perhaps contributed to the disrepute in which many writers hold the law
of Pareto; but it is clearly unfair to blame a formal expression for the
complications made inevitable by the data which it represents. If $2 < \alpha < 3$,
second moments exist, but concepts based upon third and fourth moments, such
as Pearson's measures of skewness and of kurtosis, are meaningless.

We are sure that, from the practical viewpoint, these diffi-
culties will eventually be solved. However, as of today, they are so se-
vere as to even require a re-examination of the meaning of the popular but

vague concept of "a structure." It is indeed a truism for the working
scientist, especially in fields where actual experimentation is impossible,
that the major danger of his trade is the possibility of confusion between
patterns that can only be used for "historical" description of his records,
and those that are also useful for forecasting some aspects of the future.
In particular, as we have seen, modern inference theory has taught us
always to list both the accepted and the rejected possibilities, and the
scientists' major problem is frequently to determine whether a conjectured
"relation" is significant with respect to what may be generally called
"spontaneous activity," which is the resultant of all the influences that
one cannot or does not want to control in the problem at hand, and which
is conveniently described with the help of various stochastic models. A
useful vocabulary considers the search for laws as a kind of extraction
and identification of a "signal" in the presence of "noise."

It is not enough however that all the members of a cultural group
agree upon the patterns that they read into a historical record. Indeed,
although there is unanimity in the interpretation of certain of Dr. Rohr-
schach's inkblots, they have no significance from the viewpoint of science
as a system of predictions. Broadly speaking, a pattern is scientifically
significant and is felt to have chances of being repeated, only if in
some sense its"likelihood" of having occurred by chance is very small. This
kind of significance is obviously to be assessed with the help of the tools
of statistics; unfortunately, those have been mostly designed to deal with
Gaussian alternatives and, when the chance alternative is Paretian, they
are not conservative enough by far. We believe that one will be able to
go around this difficulty, but, whenever one works in a field where the back-
ground noise is Paretian, one must begin by taking an accurate measure of
the weight of the burden of proof that one faces, and which is closer to
that of history and autobiography than to that of physics.

The same thought can be presented in more optimistic sounding words,
by saying that if a "mere chance" can so readily be confused with a causal
structure, it is itself entitled to the same noble designation, rather than
the less high-sounding term "noise." That is, "noise" may perhaps be re-
served for the Gaussian error terms, or its binomial or Poisson kins, which
are   indeed   universally disliked as sources of nuisance,
but are seldom respected as sources of anything interesting-looking.

The situation is made worse by the fact that, in models known to be very structured (e.g., to be autoregressive) with a Paretian noise, one should expect the generated paths to be much more influenced by the noise, and much less by the structure, than is the case in the Gaussian case-- where noise is already very influential. We hope to develop this point in later editions of reference 11 .

The association between the law of Pareto and "interesting patterns" is nowhere more striking than in the outcome of accumulated tosses of a coin. Indeed, the following fact is examined in the later parts of most good books on probability: suppose that we break into the game of tossing a fair coin, which "Peter" and "Paul" have been playing since sometime in the early eighteenth century. Whenever the coin falls on "heads," Peter wins a dollar (or perhaps rather a thaler); whenever the coin falls on tails, Paul wins, and let T designate the time it takes for Peter and Paul's fortunes to return to the state they were in at the moment when we broke in. For large values t of T, one has the relation:

$$\left\{ \begin{array}{l} \text{Probability that the fortunes return to their initial} \\ \text{state after a time greater than } t = (\text{constant}) \ t^{-1/2}, \end{array} \right.$$

which is the law of Pareto of exponent 1/2.

However, it is notorious that gamblers see an enormous amount of in-teresting detail in the past records of <u>accumulated</u> coin tossing gains (even more than in the separate results of tossing a coin), and that they are prepared to risk their fortunes on the proposition that these details are not due to mere chance. Similar phenomena ought to be expected when-ever the law of Pareto applies: that is, the stochastic models associated with those phenomena can well dispense with any kind of built-in causal structure, and yet generate paths in which the unskilled or the skilled eye equally well distinguishes details that are usually associated with causal relations. Similar details would be so unlikely in the path generated by a Gaussian process, that they would surely be considered as significant for forecasting. But this is not so in the Paretian case: there, from the viewpoint of prediction, those structures should be considered as being <u>perceptual illusions</u>: they are in the observer's current records and in his brain, but not in the mechanism that has generated these records and that will generate the future events.

Bearing in mind the existence of such models, let us suppose that we

have to infer a process from the data. We believe that, in many cases, a non-structured Paretian universe is capable of accounting so well for the observations, that it will be extremely difficult at best to choose between alternative models, one of which consciously imbeds some causal relations, while the other has no structure other than stochastic.

A student's belief in the existence of "genuine" structures will therefore be challengeable only with the greatest difficulty; conversely, in order to communicate such a belief to others, with the standards of creibility current in physical science, one will need much more than the tests of significance that some social scientists shrug off at the end of a discussion. Such a situation will-- as we said-- require a drastic sharpening of the distinction between patterns that-- whichever the scholar's diligence-- can only be useful for historical purposes, and those usable for forecasting the future.

The question we have in mind can be well illustrated by the problem of the significance of "cycles." With the help either of many charts or of the most sophisticated methods of Fourier analysis, it is comparatively easy to show that almost any record of the past is made up of some combination of swings. But the same is also true for a wide variety of artificial series generated by random processes with no built-in cyclic behavior whatsoever, and it is known that, however great their skill, cycle researchers seldom risk firm short-term forecasts. Could we then ask, using Keynes's terms, How far are these curves...meant to be no more than a piece of historical curve-fitting and description, and how far do they make inductive claims with reference to the future as well as the past?

It may also be noted that, because of the invariance of the law of Pareto with respect to various transformations, one cannot hope that a simple way out will be provided by arguing that only the genuine structures will be apparent to all observers. That is, the only criterion of trustworthiness is replicability in time. This again may not be a straightforward matter, because in an important respect the models of Paretian spontaneous activity diverge from the standards of "operationalism" suggested by philosophers. Indeed, in order to explain by mere chance any given set of phenomena, it will be necessary to imbed them in a universe that also contains such a fantastic number of other possibilities, that billions of years may be necessary to run through all of them. Hence, within

our lifetime, any given configuration will occur at most once and one could hardly at all define a probability for them on the basis of sample frequency. This conceptual difficulty is of course common knowledge among physicists and it is to be regretted that the philosophical discussions of the foundations of probability so seldom investigate this point. In a way, the physicists' models freely indulge in practices that for the historian are mortal sins: to rewrite history as it would have been, if Cleopatra's nose had a different shape. Our sins are even worse than the physicists', because their contrafactual histories turn out after all to be all very close to some kind of a "norm," a property which our models certainly do not possess.

We think some examples are in order here, although this section is already too long by far. We shall limit ourself to two re-interpretations of the coin-tossing record plotted on Figure   4.

First of all, forgetting the origin of that figure, let us imagine that it is a geographical cross-section of a new part of the world, in which all the regions below the bold horizontal line are under water. Let us also imagine that this chart was just brought home by an explorer (we found that most observers have no great difficulty in indulging in such a fling of the imagination), and that our problem is to decide whether it was due to cause or to chance. The naive defense will readily resort to the Highest Cause, using our graph as fresh evidence that God created Heaven and the Earth, using the same template for all the Earth, and that He also created the Verb, in which such concepts as a continent, an ocean, an island, an archipelago or a lake are precisely adapted to the shape of the Earth. Against this, the   Devil's Advocate will have no difficulty in arguing that the Earth is a creation of blind chance, and that the possibility of using such convenient terms as "continent" and "island" just reflects the chance fact that the areas above water happen to be very short or very lonp very often, and to be unexpectedly seldom of average length.

The preceding example is not as fictitious as it may seem, because the distribution of the sizes of actual island is precisely Paretian. Hence, our hypothetical debate emphasizes the two extreme outlooks realistically, even though-- the Earth having been presumably entirely explored-- no actual prediction is involved in the choice between the interpretation· of archipelagoes as "real" or as creations of the mind of the weary mariner.

Another example, also chosen for its lack of <u>direct</u> economic inter-
pretation: the problem of clusters of errors on telephone circuits. Sup-
pose that a telephone line is only used to transmit either dots or dashes,
which may be distorted in transmission to the point of being mistaken for
each other. It is clear-- according again to the defender of a search
for Causes-- that whenever an electrician touches the line, one should ex-
pect to observe a small cluster of such errors. Since moreover a screw-
driver touches the line many times during a single repair job, one should
expect to see clusters of clusters of errors, and even clusters of third
order and higher.

Actual records of the moments when errors occurred do indeed exhi-
bit such clusters, with long periods of flawless transmission in between.
A good idea of the distribution of the errors is, for example, provided by
the sequence of points where the twice-used graph of Figure   crosses the
line that used to represent sea-level. According to the searcher for
Causes, the precise study of such past records will make it possible to
better predict where new errors will occur and to minimize their effects.

On the other hand, precisely because of the origin of Figure 2 , the
Devil's Advocate can again point out that those beautiful hierarchies of
degrees of clustering can very well be due to "mere chance," devoid of any
memory and hence entirely useless for purposes of prediction.

Similar critical roles can very well be played in many other con-
texts, and we think that it is mandatory that somebody play them in every
important problem, without forgetting that the Devil's Advocate must always
be on the side of the Angels. An interesting example of stable truce be-
tween structure and chance is provided by the study of language and of dis-
course, where the traditional kind of structure is represented by grammar
and-- as one should expect by now-- the chance mechanism is akin to the
law of Pareto.(15).

7. Two-tailed Paretian variables and multi-dimensional stable Paretian laws.

We have up to now followed tradition by associating the law of Pareto
with essentially positive random variables, with a single long tail, so
that their central portion is necessarily quite skew. However, we have
discovered important examples in economics of distributions having <u>two</u>
Paretian tails (the most striking example refers to relative price changes

of sensitive speculative commodities). The argument of invariance under
maximization cannot be extended to that case. But invariance under mix-
ture simply leads to the combination of a Paretian distribution applying
to all positive u and of another applying to all negative u. As to in-
variance under aggregation, it is satisfied by every of the "stable" ran-
dom variables, which are constructed as the sum or the difference between
two arbitrarily weighted "positive" stable variables of the kind studied
earlier in this paper. In particular, stable variables can be symmetric;
the Cauchy distribution provides a prime example. But their study depends
very little upon the actual degree of skewness; hence, the asymmetry of the
usual Paretian variables is less crucial than the length of their single
tail.

Another remarkable property of the stable distributions is that, like
the Gaussian, they have intrinsic extensions to the multi-variate case,
other than the degenerate case of independent coordinates. Very few other
distributions share this property, and the reason for this is intimately
related to the role of stable distributions in linear models: it is indeed
possible to characterize the multi-variate stable distributions as being
those for which the distribution of every linear combination of the coordin-
ates is a scalar stable variable. This property is essential in the study
of multi-dimensional economic quantities, as well as in the investigation
of the dependence between successive values of a one-dimensional quantity
such as income (see reference 8).

8. Conclusion concerning the role of Pareto's law in economics and establishment of a link with the physical sciences.

Our arguments show that there is strong pragmatic reason to begin the
study of economic distributions and time-series by those that satisfy the
law of Pareto. Since this category includes prices (reference 11), firm
sizes (reference 13), and incomes (references 7 , 8 and 9), the study
of Paretian law acquires a fundamental importance in economic statistics.

Similarly, the example of the distribution of city sizes stresses the
importance of the law of Pareto in sociology (reference 10). Finally,
we have strong indications of its importance in psychology. (We shall not
even attempt to outbid George Kingsley Zipf in listing all the Paretian
phenomena of which we are aware; their number seems to increase all the time.)

However, it is impossible to postpone "explanation" forever. If indeed a grand Economic System is only based upon aggregation, choice and mixture, one can prove that, in order that it be Paretian, it **must** be triggered somewhere by a Paretian "initial" condition. That is, however useful the method of invariants may be, it is true that it somewhat begs the question, and that the basic mystery cannot be solved by pushing it around. Indeed, if it were true, in accordance with "conventional wisdom", that physical phenomena are characterized by the law of Gauss, and social phenomena by that of Pareto, we may eventually have to explain the latter by some of the "micro-scopic" economic models, such as the "principle" of random proportionate effect (reference 14) which we prefer to de-emphasize in our approach.

We claim, however, that such **need not** be the case. Quite on the contrary, the physical world is full of Paretian phenomena which one can easily visualize as playing the role of the "triggers" that cause the economic system to be also Paretian. We found for example (reference 12) that single-tailed Paretian distributions, with trustworthy values for α , represent the statistical distributions of a variety of natural resources, which are surely not influenced by the structure of society, and by weather, which is barely influenced by man, as yet. Such is the case of the areas of oil fields and their total capacities (i.e., the sums of the total production and of the currently estimated capacity); the same is true for the valuations of certain gold, uranium and pyrite mines in South Africa, for at least some levels of rivers, and for a host of similar data related to weather-- some of which, such as hail, have a direct influence on important risk phenomena, namely the insurance against hail damage.

If our purpose were to contribute to "geo-statistics," we should of course examine the degree of generality of our claim. But, for the purpose of a study of economic time-series, it will be quite sufficient to note that a Paretian Grand Economic System **can very well** be triggered by statistical features of the physical world. For example, natural resources and weather influence prices, which in turn influence incomes. (Since the Systems to which we refer are spatio-temporal, there is nothing disturbing in our association of economic **time**-series with geological and geographical **spatial** distributions.)

We shall not attempt to say anything about the actual triggering

mechanism, since we doubt that a unique link can be found between the social and the physical worlds. After all, quite divergent values of Pareto's alpha are encountered in both so that the overall Grand System cannot possibly be based only upon transformations by linear aggregation, choice and mixture.

Let us also point out that, even if one finds simple models for the various occurrences of Pareto's law in geomorphy, many aspects of this general problem will be accessible to our "phenomological analysis" and for many purposes they should be so treated. Moreover, until models become available, this is the only open alternative.

We wish finally to point out that the Paretian phenomena of physics have also turned out to include some that are devoid of direct relation with economics. For example, a three-dimensional stable law occurs in the theory of Newtonian attraction (reference 7). Moreover, the distribution of the energies of the primary cosmic rays has long been known to follow a law which happens to be identical to that of Pareto with the exponent 1.8 (as a matter of fact, Enrico Fermi's study of this problem happens to include an unlikely, but rather neat generation for the Pareto distribution; see reference 14). The same holds for meteorite energies and is important for ionospheric scatter telecommunications. Also, the intervals between successive errors of transmission on telephone circuits happen to be Paretian with a very small exponent, the value of which depends upon the physical properties of the circuit (see reference 1), as discussed in section 6. This example-- combined with the problem of the areas of islands and lakes also investigated in section 6-- suggests that many of the Paretian phenomena encountered in practice may be related to "accumulative" processes similar to those encountered in coin tossing.

In any event, all the examples of a Paretian behavior show that statisticians will have to pay special attention to distributions without population moments.

9. An examination of Frederick Macaulay's criticism of the law of Pareto.

Finding so many reasons for considering the law of Pareto as being one of the most important of all probability distributions, we were of course permanently surprised by the "neglect and even contempt" to which we referred in the first sentence of this paper. We eventually found that this

attitude had deep roots not only in the apparent lack of theoretical mo-
tivation for that law, but also in several seemingly "definitive" criti-
cisms, and we would like to analyze two often quoted adverse analyses.

We shall begin with F. R. Macaulay's (reference 6). We found this
essay most impressive and-- even though we obviously disagree with its
conclusions-- we recommend very strongly that it continue to be read. It
has indeed fully disposed of any possible claims concerning the invariance
of Pareto's exponent from year to year and from country to country, and con-
cerning the relevance of the law of Pareto to the description of small in-
comes or of incomes of the lower-paid professional categories. Macaulay is
also very convincing concerning Paretian distributions with a high exponent
(see section 5; his conclusions on this account were independently reached--
much later-- in 9).

We definitely believe, however, that his main point is not well-taken
and that his strictures against what is called "mere curve-fitting" have
been very harmful. Indeed, his ideals of a proper mathematical description
are not followed in any science we know of, and they have materially contri-
buted to the excessive reliance of statistical economics upòn Gaussian or
lognormal "null hypotheses," which are patently wrong in most cases, or
upon non-parametric methods, which by definition cannot possibly tell very
much about any specific situation. One should of course only use curve-
fitting for what it is worth, but not for any less.

For example, Macaulay points out that an excellent fit of the cumulated
expression $Pr(U \leqslant u)$ ("global" limit theorems)only As a result, one had better
avoid inferences from densities; if one cannot avoid them, one should not
expect them to be very good.

But Macaulay is even more severe and he finds that the empirical
curves do not zigzag around the simple Paretian interpolate, but rather
cross it systematically a few times. The fact that this observation was
used to reject the law of Pareto outright illustrates a basic difference be-
tween the outlooks of the careful economists and of the careless physicists:
when the law of Boyle was similarly found to differ from facts, the physi-
cists simply invented the concept of a "perfect gas," that is, a body that
follows perfectly Boyle's law. Naturally, perfect-gas approximations are
not even considered in some problems (for example, such bodies never cease
to be gases, and-- whichever the temperature-- they cannot become liquid).

However, the perfect-gas approximation is adequate in many cases, and it is
so simple that one cannot afford not to consider it first. Similarly,
Pareto-law approximations should not even be considered in some problems
(for example, those relative to low incomes); but one cannot afford not to
consider them first in other investigations.

Macaulay's criticism of the law of Pareto may therefore be summarized
from our viewpoint by saying that it only endorses the "weak" forms of this
law with which we had occasion to work. In many cases, however, we think
that it is legitimate to take more seriously certain Paretian kins, such as
the stable distributions.

We fell less well disposed towards other critiques of Pareto's law,
such as Dwight B. Yntema's (reference 16). This work happens indeed to be
a call for the measurement of inequality by various expressions based on
sample moments, rather than by Pareto's exponent $\propto$. We agree of course
that Pareto's exponent is insufficient, as long as the concept of "inequali-
ty" is defined so as to involve medium and small incomes. But, if the con-
cept of "inequality" is defined so as to involve large incomes, we have shown
that the sample moments are nonsensical. There is as yet no common ground
to compare the indices of different kinds, so that Yntema's evidence is ir-
relevant to the validity of Pareto's law.

<center>Footnotes</center>

(1) This is of course only an analytic way of saying that, if one plots
the logarithm of the number of incomes greater than u, as a function of
the logarithm of u, one should obtain a curve that for large u becomes a
stright line sloping down to the right with an absolute slope equal to α .

(2) That is, the method of invariants used by physicists is a somewhat
different procedure. For example, the classical "principle of relativity"
was not introduced to "explain" any complicated empirical law such as
that of Pareto. For the stress upon the nuances between different methods
of invariants we are indebted to Harrison White.

(3) <u>Mixture</u>. It is easy to see that one has
$$\Pr(U_w > u) \sim \sum C_n p_n u_n^{-\alpha} = C_w u^{-\alpha} \text{ with } C_w = \sum p_n C_n. \quad \text{QED}$$

(4) <u>Maximizing choice</u>. In order that $U_M \leqslant u$, it is clearly both necessary
and sufficient that $U_n \leqslant u$ <u>for every</u> n. Hence, $\Pr(U_M \leqslant u) = \prod \Pr(U_n \leqslant u)$.
It follows that one has:
$$\Pr(U_M > u) = 1 - \Pr(U_M \leqslant u) \sim 1 - \prod (1 - C_n u^{-\alpha}) \sim \sum C_n u^{-\alpha} = C_M U^{-\alpha}, \quad \text{QED}.$$

(5) <u>Aggregation.</u> Here the argument is more involved, and we prefer to
suggest to the reader to look up the proof in reference 7 .

(6) Let U be characterized by its distribution function $F(u) = \Pr(U \leqslant u)$
and by its generating function G(s), which is the Laplace transform of F(u);
$G(s) = \int_{-\infty}^{\infty} \exp(-u s) dF(u)$. (This limits our argument to laws for which dF
is so small for $u < 0$ that G converges.) Then, one can begin by writing the
following conditions, which are respectively necessary for the various types
of invariance-- up-to-scale.

<u>Weighted Mixture</u>. It is necessary that stability hold for equal p_n.
Thus, it is in particular necessary that the function F satisfy the condi-
tion that
$$\frac{1}{N} \sum F(u/a_n) = F(u/a_w)$$

<u>Maximization</u>. Now, it is necessary that $F(u/a_M) = \prod F(u/a_n)$; in other words, one must have:

$$\sum \log F(U/a_n) = \log F(u/a_M).$$

<u>Aggregation</u>. This requires that $G(a_a s) = \prod G(a_n s)$; in other words, one must have:

$$\sum \log G(a_n s) = \log G(a_a s).$$

It turns our therefore that the three types of invariance lead to formerly almost identical equations, although they refer to different functions, respectively F_w, $\log F_M$ and $\log G_A(s)$. The general solutions must therefore respectively take the forms $F_w(u) = C' - Cu^{-\alpha}$; $F_M(u) = \exp(-Cu^{-\alpha})$ and $G_A(s) = \exp(-Cs^{-\alpha})$. One also easily verifies that $a_M^\alpha = a_A^\alpha = \sum a_n^\alpha$, $a_w^\alpha = \frac{1}{N} \sum a_n^\alpha$.

Now, we shall show that the above necessary conditions are actually not sufficient, and that additional requirements must be imposed upon C', C and α.

<u>Maximization</u>. The distribution function of a random variable must be non-decreasing and such that $F_M(\infty) = 1$. This requires that $C > 0$ and $\alpha > 0$, which leaves us with the laws $F_M(u) = \exp(-Cu^{-\alpha})$.

<u>Mixture</u>. In order that $F_w(u)$ be non-decreasing and such that $F_w(\infty) = 1$, it is now necessary that $C' = 1$, $\alpha > 0$ and $C > 0$.

<u>Aggregation</u>. In order that $G_A(s)$ be a generating function, one can show that it is necessary that $0 < \alpha < 1$ with $C < 0$, or $1 < \alpha \leq 2$ with $C > 0$.

COMMENTED REFERENCES

[1]   JAY M. BERGER and B. MANDELBROT, "A new model for the cluster-
      ing of errors on telephone lines", IBM Journal of Research
      and Development, Vol. 7, July 1963. (A Paretian problem
      indirectly related to economics, but very simple in form.)

[2]   W. FELLER, An Introduction to Probability Theory and its Appli-
      cations, Vol. 1, 2nd Edition, New York, John Wiley, 1957.
      (The standard treatise on coin-tossing, and some related matters.)

[3]   G. L. GERSTEIN and B. MANDELBROT, "Random Walk models of the
      spontaneous activity of single neurons", submitted to The
      Biophysical Journal

[4]   B. V. GNEDENKO and A. N. KOLMOGOROFF, Limit distributions for
      sums of independent random variables, English translation by
      K.L. Chung, Reading, Mass., Addison-Wesley Press, 1954.
      (This excellent textbook reproduces most of the substance -
      but not the inimitable flavor - of the following two great
      books: Paul LEVY, Calcul des probabilités, Paris, Gauthier-
      Villars, 1925 and Paul LEVY, Théorie de l'addition des
      variables aléatoires, Paris, Gauthier-Villars, 1937 and 1954.)

[5]   E. GUMBEL, Statistics of Extremes, New York, Columbia University
      Press, 1958. (The standard textbook on maximization, including
      the theory originating it: M. FRECHET, "Sur la loi de proba-
      bilité de l'écart maximum", Annales de la Société Polonaise
      de Mathématiques (Cracovie), Vol. 6, 1927, pp. 93...)

[6]   F. MACAULAY, "Pareto's law: and the general problem of mathe-
      matically describing the frequency distribution of income",
      Income in the United States, its amount and distribution,
      1909-1919, (Chapter 23 of Volume II) New York, National
      Bureau of Economic Research, 1922.

[7]   B. MANDELBROT, "The Pareto-Lévy Law and the distribution of
      income", International Economic Review, Vol. 1, 1960,
      pp. 79-106, as amended in "The Stable Paretian income dis-
      tribution, when the apparent exponent is near two", Inter-
      national Economic Review, Vol.4, January 1963.

[8] B. MANDELBROT, "Stable Paretian random functions and the
     multiplicative variation of income", Econometrica, Vol. 29,
     1961, pp. 517-543. (We think that one should postpone to the
     end these more involved considerations upon income variation.)

[9] B. MANDELBROT, "Paretian distributions and income maximization ",
     Quarterly Journal of Economics, Vol. 76, 1962, pp. 57-85.
     (In this paper, the operation of maximization is extended to
     the case of non-independent U_n.)

[10] B. MANDELBROT, "Very long-tailed probability distributions and
     the empirical distribution of city sizes", to appear in
     Mathematical Approaches to the Social Sciences, edited by
     F. Massarik and P. Ratoosh, published by Richard D. Irwin, 1963.

[11] B. MANDELBROT, "The variation of certain speculative prices",
     IBM Research Report NC-87, March 1962 (out of print; to
     appear soon as a separate monograph.)

[12] B. MANDELBROT, "Statistics of natural resources and the law of
     Pareto", IBM Research Note NC-146, June 29, 1962 (soon to
     be superseded by a more complete account.)

[13] B. MANDELBROT, "Oligopoly, merger and the Paretian size distri-
     bution of firms", IBM Research Report (in print)

[14] B. MANDELBROT, "Survey of the Growth and Diffusion Models of
     the law of Pareto", IBM Research Report (in print)

[15] B. MANDELBROT, "On the theory of word frequencies and on
     related markovian models of discourse", pp. 190-219 of
     Structure of language and its mathematical aspects, edited
     by Roman Jakobson, and published by the American Mathematical
     Society (1960.) B. MANDELBROT, "Linguistique statistique
     macroscopique", pp.1-80 of Logique, language et théorie de
     l'information, by L. Apostel, B. Mandelbrot et A. Morf,
     published by Presses Universitaires de France (1957) (out
     of print; a second edition is in preparation.)

[16] D. B. YNTEMA, "Measures of inequality in the personal distri-
     bution of wealth or income", Journal of the American Statistical
     Association, Vol. 28, December 1933, pp. 423-433.

[17]   G. K. ZIPF, <u>National Unity and Disunity</u>, Bloomington, Indiana, 1941.

[18]   G. K. ZIPF, <u>Human Behavior and the Principle of Least-Effort</u>,
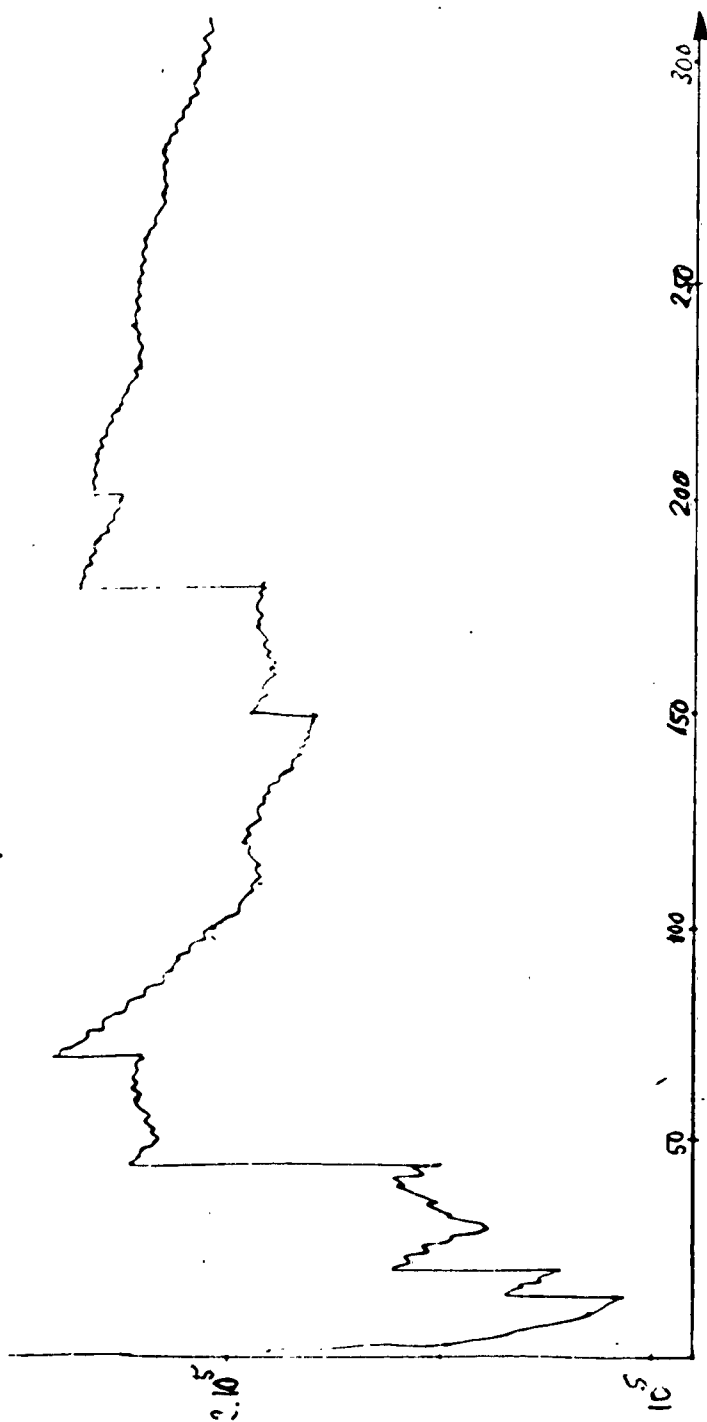   Cambridge, Mass., Addison-Wesley Press, 1949.

FIGURE 1. Example of a record of successive values of the sample first moment. when the sample values are drawn from a Paretian population with an alpha close to 1, so that the population moment is very large or infinite.
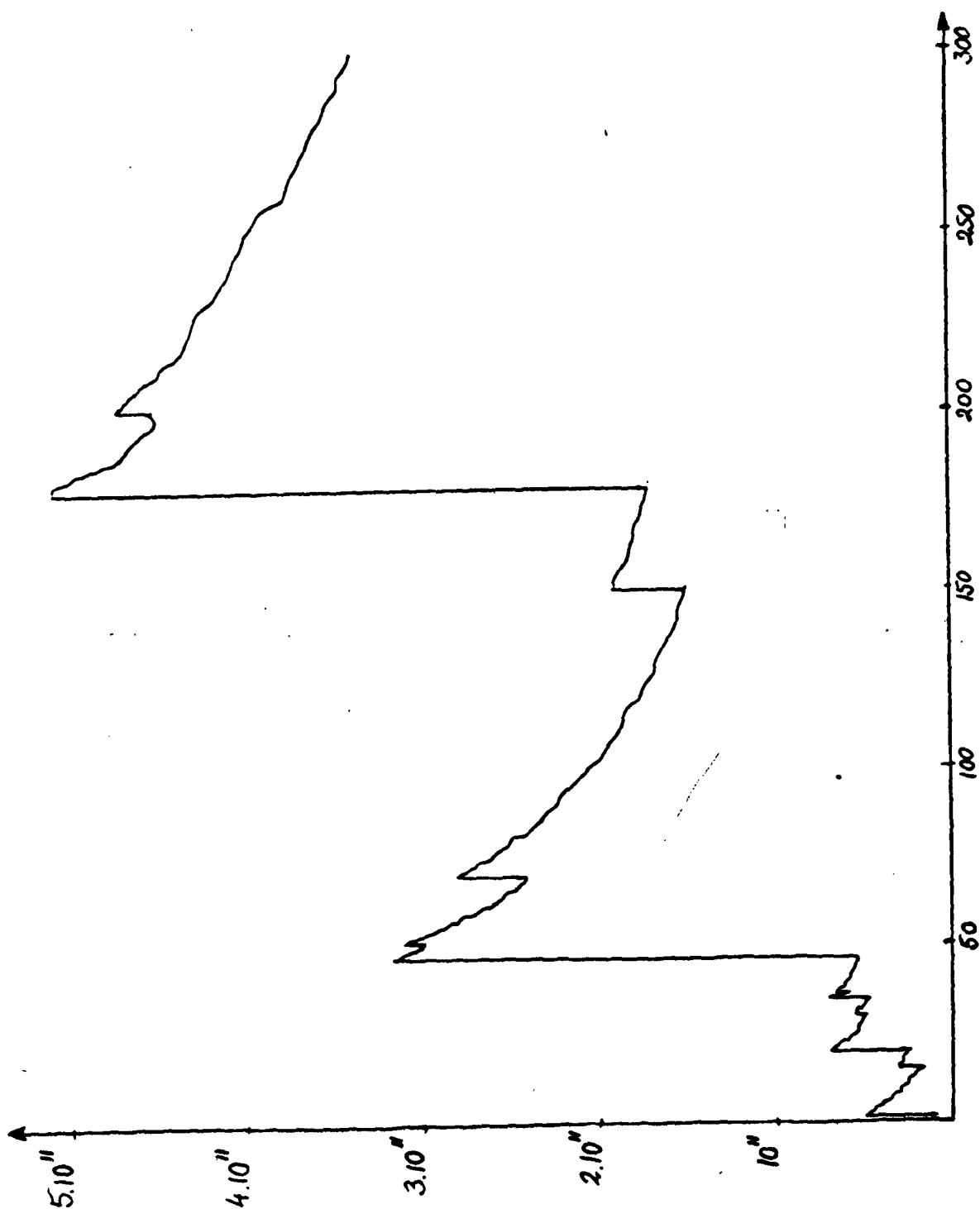
FIGURE 2. Example of a record of successive values of the sample second
moment, when the sample values are drawn from a Paretian population
with an alpha close to 1/2, so that the population moment is surely infinite.
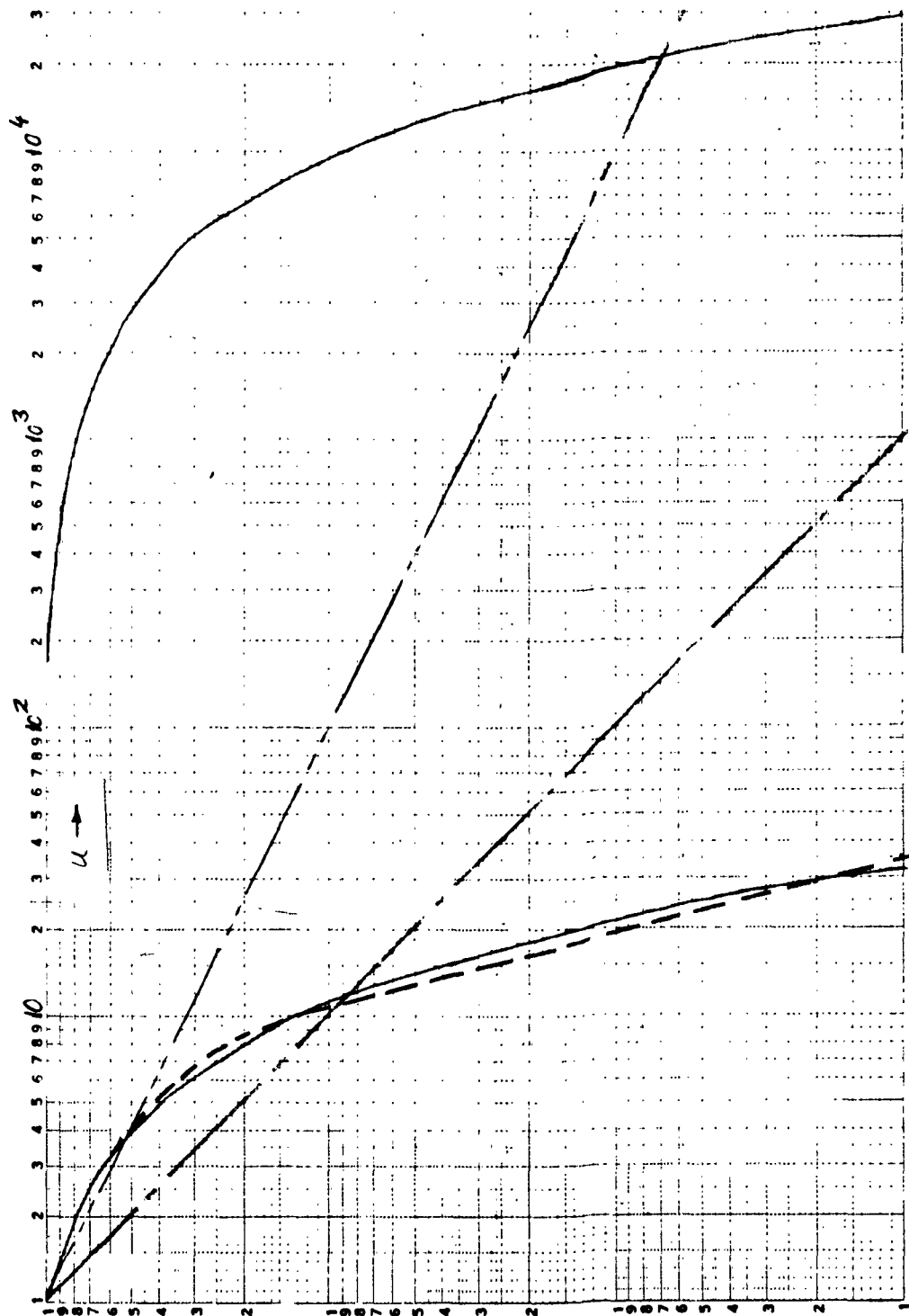
FIGURE 3. A doubly logarithmic (Paretian) graph, on which we have plotted: A)
Two exponential distributions (very curved solid lines), having very different
means. B) Two distributions satisfying Pareto's law all through (from u = 1
on) and having the exponents 1/2 and 1. C)A distribution having asymptotically
a Paretian exponent of 4. It is obvious that the last distribution can readily
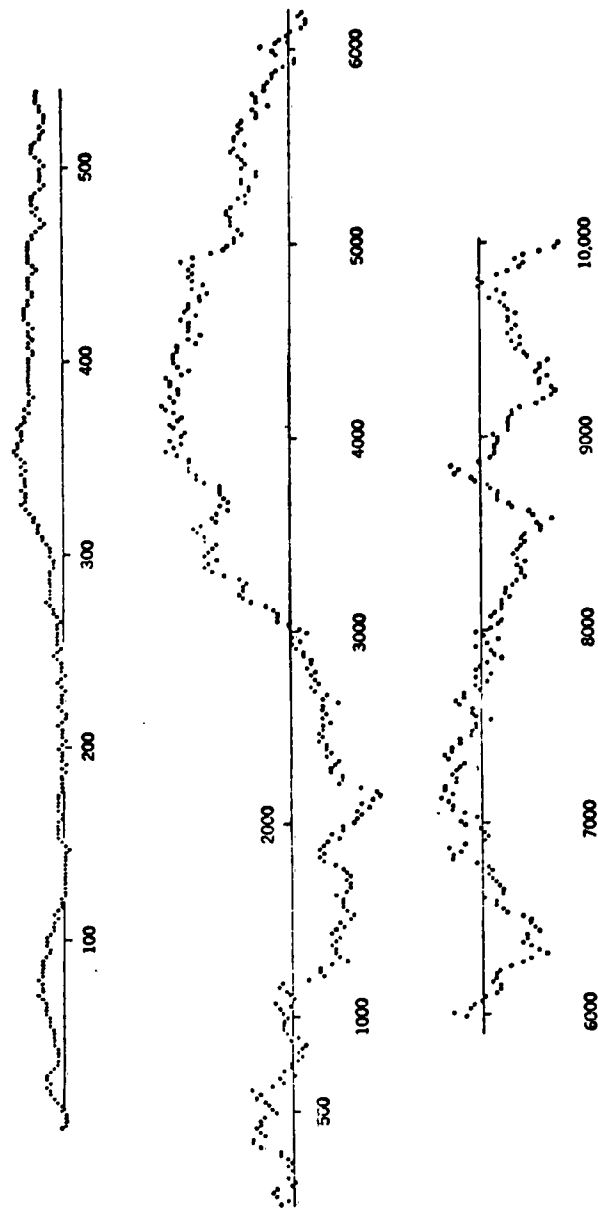be confused with the exponential but that small alpha exponents are reliable.

FIGURE 4. Record of Paul's winnings in a coin-tossing game, played with a fair coin. Zero-crossings seem to be strongly clustered, although the intervals between crossings are obviously statistically independent. This Figure is reproduced from Feller (reference 2.)

In order to appreciate fully the extent of apparent clustering in this Figure, one must note that the unit of time used on the second and third lines equals 20 plays. Hence, the second and third lines lack detail and each of the corresponding zero-crossings is actually a cluster or a cluster of clusters. For example, the details of the clusters around time 200 can be clearly read on line 1, which uses a unit of time equal to 2.